

Fourier domain preconditioned conjugate gradient algorithm for atmospheric tomography

Qiang Yang, Curtis R. Vogel, and Brent L. Ellerbroek

By “atmospheric tomography” we mean the estimation of a layered atmospheric turbulence profile from measurements of the pupil-plane phase (or phase gradients) corresponding to several different guide star directions. We introduce what we believe to be a new Fourier domain preconditioned conjugate gradient (FD-PCG) algorithm for atmospheric tomography, and we compare its performance against an existing multigrid preconditioned conjugate gradient (MG-PCG) approach. Numerical results indicate that on conventional serial computers, FD-PCG is as accurate and robust as MG-PCG, but it is from one to two orders of magnitude faster for atmospheric tomography on 30 m class telescopes. Simulations are carried out for both natural guide stars and for a combination of finite-altitude laser guide stars and natural guide stars to resolve tip-tilt uncertainty. © 2006 Optical Society of America

OCIS code: 010.1080

1. Introduction

Adaptive optical (AO) systems now under development for future extremely large telescopes will employ multiple guide stars and tomographic wavefront reconstruction algorithms to estimate the full three-dimensional profile of atmospheric turbulence.^{1–4} The formal theory of atmospheric tomography is by now well understood,^{5–11} and is fundamentally a generalization of the so-called minimum variance wavefront reconstruction algorithms developed previously for the case of narrow-field AO system with a single natural or laser guide star.^{12,13} Work is now in progress toward computationally efficient tomographic reconstruction algorithms suitable for use with wavefront sensors with on the order of 10^4 subapertures and deformable mirrors with similar numbers of actuators.

Two general approaches have been considered to date: conjugate gradient iterations with multigrid preconditioning^{14–16} (MG-PCG), and Fourier domain methods that exploit the fact that the basic opera-

tions of wavefront propagation and wavefront gradient sensing may be modeled as spatial filters in the Fourier domain.^{17–19} Both methods have been applied to very-high-order reconstruction problems derived from atmospheric tomography, but neither can be considered entirely successful at this time. The efficiency of MG-PCG scales with $N^{3/2}$ (where N is the order of the tomographic reconstruction problem), while finite apertures degrade the optimality and/or computational efficiency of Fourier domain algorithms. This paper investigates whether the various strengths of the two approaches may be merged into a single tomographic reconstruction algorithm.

We propose to use a conjugate gradient reconstruction algorithm with a Fourier domain preconditioner (FD-PCG) instead of multigrid sweeps. As already noted above, Fourier domain methods are attractive because the wavefront sensing operator that must be inverted to perform atmospheric tomography is nearly a spatial filtering operation. This operator is the product of three terms: wavefront propagation from the guide star through the atmosphere, differentiation of the pupil-plane phase by a wavefront gradient sensor, and masking by the clear aperture of the telescope. The first two operators are shift invariant, and correspond to spatial filters in the Fourier domain. The final operator is multiplicative, and corresponds to a convolution. By truncating the kernel of this convolution, we obtain a sparse approximation to the (Fourier domain representation of) wavefront sensing operator that may be used as the basis for a computationally efficient preconditioner.

Qiang Yang (yang@math.montana.edu) and C. R. Vogel are with the Department of Mathematical Sciences, Montana State University, Bozeman, Montana 59717-2400. B. L. Ellerbroek is with the TMT Project Office, California Institute of Technology, Pasadena, California 91126.

Received 25 February 2005; revised 20 September 2005; accepted 15 December 2005; posted 19 December 2005 (Doc. ID 60047).

0003-6935/06/215281-13\$15.00/0

© 2006 Optical Society of America

Simulation results confirm that this preconditioner is not only computationally efficient, but also provides essentially the same level of performance as the earlier MG-PCG algorithm in terms of convergence rates and asymptotic residual errors. But each iteration of FD-PCG is from 10 to 100 times faster than MG-PCG for typical simulations of atmospheric tomography on 30 m class telescopes. By using the same low-rank matrix manipulations previously devised for the MG-PCG approach, we are also able to apply the FD-PCG algorithm to simulations modeling laser guide star (LGS) position uncertainty and natural guide star (NGS) tip-tilt sensors. We believe that this represents an advance over existing Fourier domain techniques.

The remainder of this paper is organized as follows. Section 2 is a brief review of minimum variance wavefront reconstruction as applied to atmospheric tomography. Section 3 describes the FD-PCG algorithm for the case of natural guide stars, and Section 4 extends the basic method to consider guide stars at finite range, guide stars with tip-tilt uncertainty, and tip-tilt natural guide stars. Section 5 presents sample numerical results for both the FD-PCG and MG-PCG algorithms, and compares their performance. Section 6 is a brief summary.

2. Minimum Variance Turbulence Profile Estimation

In this section we review the minimum variance approach to atmospheric tomography. Open-loop wavefront sensor (WFS) measurements are modeled as

$$s = G\psi + \eta, \quad (1)$$

where ψ represents the turbulence profile for a layered atmosphere, G is a linear operator that maps turbulence profile to idealized sensor measurements, and η is the WFS measurement noise.

The minimum variance turbulence profile estimate is given by

$$\psi_{\text{MV}} = E_{\text{MV}} s, \quad (2)$$

where the estimation matrix E_{MV} minimizes

$$J(E) = \langle \|Es - \psi\|^2 \rangle = \text{Trace} \langle (Es - \psi)(Es - \psi)^T \rangle. \quad (3)$$

Here $\langle \dots \rangle$ represents the ensemble average. By assuming that $\langle \psi \rangle = 0$, $\langle \eta \rangle = 0$, and that ψ and η are independent, we obtain^{14,20}

$$E_{\text{MV}} = (G^T C_\eta^{-1} G + C_\psi^{-1})^{-1} G^T C_\eta^{-1}, \quad (4)$$

where $C_\eta = \langle \eta \eta^T \rangle$ is the noise covariance matrix and $C_\psi = \langle \psi \psi^T \rangle$ is the atmospheric turbulence covariance matrix. Plugging Eq. (4) into Eq. (2) yields

$$\psi_{\text{MV}} = (G^T C_\eta^{-1} G + C_\psi^{-1})^{-1} G^T C_\eta^{-1} s. \quad (5)$$

This solution could also be obtained using maximum *a posteriori* (MAP) estimation²¹ if one assumes Gaussian statistics. If we let x be a discrete vector representation for ψ and set

$$A = G^T C_\eta^{-1} G + C_\psi^{-1}, \quad (6)$$

$$b = G^T C_\eta^{-1} s, \quad (7)$$

then our problem is to solve the symmetric and positive definite (SPD) linear system $Ax = b$.

3. Fourier Domain Preconditioning

A. Preconditioned Conjugate Gradient Method

The preconditioned conjugate gradient (PCG) method²² can be an efficient and robust scheme to iteratively solve large SPD linear systems $Ax = b$. In principle, PCG is standard conjugate gradient iteration applied to a symmetrized version of the transformed system

$$C^{-1}Ax = C^{-1}b, \quad (8)$$

where the preconditioning matrix C is also SPD. In order for PCG to be effective, matrix-vector products $C^{-1}r$ must be inexpensive to compute and $C^{-1}A$ must have a desirable eigenvalue distribution in the sense that its eigenvalues are clustered and/or have a relatively small range. The PCG algorithm can be organized in the following form:

1. Compute $r_0 = b - Ax_0$, $z_0 = C^{-1}r_0$, and $p_0 = z_0$
2. For $k = 0, 1, \dots$, until convergence Do
3. $q_k = Ap_k$
4. $\alpha_k = (r_k, z_k) / (q_k, p_k)$
5. $x_{k+1} = x_k + \alpha_k p_k$
6. $r_{k+1} = r_k - \alpha_k q_k$
7. $z_{k+1} = C^{-1}r_{k+1}$
8. $\beta_k = (r_{k+1}, z_{k+1}) / (r_k, z_k)$
9. $p_{k+1} = z_{k+1} + \beta_k p_k$
10. EndDo

The notation (\dots, \dots) in steps 4 and 8 indicates inner product, or dot product, of a pair of vectors. These inner products have $\mathcal{O}(n)$ cost, where n is the system size. The scalar-vector products together with vector sums in steps 5, 6, and 9 also have $\mathcal{O}(n)$ cost. The dominant costs at each PCG iteration are the matrix-vector products involving the matrix A in step 3 and the inverse of the preconditioner C in step 7.

B. Fourier Domain Preconditioning

Let L denote the number of layers in the atmospheric turbulence profile ψ , and let the ℓ th layer be represented by a "phase screen" ψ_ℓ , defined on an $n_x \times n_y$ grid. Hence ψ can be viewed as a grid function, which maps points on an $n_x \times n_y \times L$ grid to real numbers. The corresponding vector representation x is $n \times 1$, where $n = n_x n_y L$, and its components consist of ψ

evaluated at the grid points. In the discussion to follow, we will freely interchange grid functions and vector representations. We will also interchange linear operators defined on grid functions with matrices that act on the corresponding vectors.

We consider preconditioners of the form

$$C = \mathcal{F}^{-1} \hat{C} \mathcal{F}. \quad (9)$$

Here \mathcal{F} is the $L \times L$ block diagonal matrix whose diagonal blocks are each (an $n_x n_y \times n_x n_y$ matrix representer for) the 2D Fourier transform. Each block corresponds to a layer in a layer-wise decomposition $[\psi_1, \dots, \psi_L]$ of ψ . In addition, \hat{C} is sparse and SPD. Hence implementation of step 7 of the PCG algorithm requires (i) the 2D Fourier transform applied to each the L blocks of r_{k+1} to obtain \hat{r} , (ii) solution of a sparse linear system $\hat{C}\hat{z} = \hat{r}$ and (iii) the inverse Fourier transform applied to the blocks of \hat{z} to obtain z_{k+1} .

Our choice for the Fourier representer \hat{C} of the preconditioner is motivated by the form of the matrix A in Eq. (6). Consider first the inverse covariance matrix C_ψ^{-1} . We assume the layers are independent, so

$$C_\psi = \text{Block Diag}\{C_{\psi_\ell}\}_{\ell=1}^L. \quad (10)$$

We assume a Kolmogorov model for turbulence at each layer, so

$$C_{\psi_\ell} = C_n^2(z_\ell) \mathcal{F}^{-1} |\kappa|^{-11/3} \mathcal{F}, \quad \ell = 1, \dots, L. \quad (11)$$

The scalar $C_n^2(z_\ell)$ quantifies the turbulence strength at layer height z_ℓ , and the grid function $|\kappa|$, which represents the magnitude of the (2D vector) wavenumber, is interpreted as a pointwise multiplication operator. Combining Eqs. (10) and (11) gives

$$C_\psi^{-1} = \mathcal{F}^{-1} \text{Block Diag}\{C_n^{-2}(z_\ell) |\kappa|^{11/3}\}_{\ell=1}^L \mathcal{F}. \quad (12)$$

Next consider the term $G^T C_\eta^{-1} G$ in Eq. (6). For simplicity, we assume geometric optics propagation and natural guide stars. Laser guide stars are considered in Section 4. In this case the pupil-plane phase that arises from light propagation through the atmosphere from a natural guide star in direction θ_j can be represented as

$$\phi_j(\mathbf{x}) = [P(\theta_j)\psi](\mathbf{x}) \stackrel{\text{def}}{=} \sum_{\ell=1}^L \psi_\ell(\mathbf{x} + z_\ell \theta_j) \quad (13)$$

$$= \sum_{\ell=1}^L \mathcal{F}^{-1} \exp(i2\pi z_\ell \kappa \cdot \theta_j) \mathcal{F} \psi. \quad (14)$$

Here \mathbf{x} denotes location of a grid point in the pupil plane, and the second equality follows from the Fourier shift theorem. Equations (13) and (14) hold

for $j = 1, \dots, N_{\text{GS}}$, where N_{GS} denotes the number of guide stars.

We assume the sensor measurements are discrete approximations to the gradient of phase evaluated on the pupil-plane grid. If we let Γ_x and Γ_y denote the x and y components of the discrete gradient operator, we can express noise-free sensor measurements as

$$s_j = \begin{bmatrix} M\Gamma_x \phi_j \\ M\Gamma_y \phi_j \end{bmatrix} = \begin{bmatrix} M\Gamma_x P(\theta_j)\psi \\ M\Gamma_y P(\theta_j)\psi \end{bmatrix}, \quad j = 1, \dots, N_{\text{GS}}. \quad (15)$$

Here M is an aperture masking operator. For simplicity, we take

$$Ms(\mathbf{x}) = \begin{cases} s(\mathbf{x}), & \mathbf{x} \in \Omega, \\ 0, & \text{otherwise,} \end{cases} \quad (16)$$

where Ω denotes the set of grid points \mathbf{x} that are within the clear aperture of the telescope.

With Fried geometry we can represent Γ_x in terms of shift operators,

$$\begin{aligned} \Gamma_x \phi &= \frac{1}{2} (S_x - 1)\phi + \frac{1}{2} (S_x - 1)S_y \phi \\ &= \frac{1}{2} (S_x - 1)(S_y + 1)\phi, \end{aligned} \quad (17)$$

where $S_x \phi(x_i, y_j) = \phi(x_{i+1}, y_j)$ and $S_y \phi(x_i, y_j) = \phi(x_i, y_{j+1})$. Similarly,

$$\Gamma_y \phi = \frac{1}{2} (S_y - 1)(S_x + 1)\phi. \quad (18)$$

By again applying the Fourier shift theorem, we can represent

$$\Gamma_x = \mathcal{F}^{-1} \hat{\Gamma}_x \mathcal{F}, \quad \Gamma_y = \mathcal{F}^{-1} \hat{\Gamma}_y \mathcal{F}, \quad (19)$$

where $\hat{\Gamma}_x$ and $\hat{\Gamma}_y$ are diagonal matrices.

Now assume the noise covariance takes the form

$$C_\eta = \sigma^2 I, \quad (20)$$

where I is the identity matrix of appropriate size. This assumption is not appropriate for laser guide stars (see Section 4). From the discussion above we obtain an $L \times L$ block representation for $G^T C_\eta^{-1} G$, with blocks

$$\begin{aligned} [G^T C_\eta^{-1} G]_{\ell m} &= \sigma^{-2} P^T(\theta_\ell) (\Gamma_x^T M^T M \Gamma_x + \Gamma_y^T M^T M \Gamma_y) \\ &\quad \times P(\theta_m) \end{aligned} \quad (21)$$

$$\begin{aligned} &= \sigma^{-2} \sum_{j=1}^{N_{\text{GS}}} P_{j\ell}^T (\Gamma_x^T M^T M \Gamma_x + \Gamma_y^T M^T M \Gamma_y) \\ &\quad \times P_{jm} \end{aligned} \quad (22)$$

$$\begin{aligned} &= \mathcal{F}^{-1} \left\{ \sigma^{-2} \sum_{j=1}^{N_{\text{GS}}} \hat{P}_{j\ell}^* (\hat{\Gamma}_x^* \hat{M}^* \hat{M} \hat{\Gamma}_x \right. \\ &\quad \left. + \hat{\Gamma}_y^* \hat{M}^* \hat{M} \hat{\Gamma}_y) \hat{P}_{jm} \right\} \mathcal{F}. \end{aligned} \quad (23)$$

Here the superscript $*$ denotes the matrix conjugate transpose. Note that $\hat{\Gamma}_x = \mathcal{F}^{-1}\Gamma_x\mathcal{F}$, $\hat{\Gamma}_y = \mathcal{F}^{-1}\Gamma_y\mathcal{F}$, and the $\hat{P}_{j\ell} = \mathcal{F}^{-1}P_{j\ell}\mathcal{F}$ all have $n_x n_y \times n_x n_y$ diagonal matrix representations. Unfortunately, $\tilde{M} = \mathcal{F}^{-1}M\mathcal{F}$ has a full matrix representation. In Subsection 3.C, we obtain sparse matrix approximations, \tilde{M} to \tilde{M} . Given such a sparse approximation, we take the Fourier representer for the preconditioner in Eq. (9) to be block $L \times L$ with blocks

$$[\hat{C}]_{\ell m} = \sigma^{-2} \sum_{j=1}^{N_{GS}} \hat{P}_{j\ell}^* (\hat{\Gamma}_x^* \tilde{M}^* \tilde{M} \hat{\Gamma}_x + \hat{\Gamma}_y^* \tilde{M}^* \tilde{M} \hat{\Gamma}_y) \hat{P}_{jm} + \delta_{\ell m} C_n^{-2}(z_\ell) |\kappa|^{11/3}, \quad (24)$$

where the Kronecker delta, $\delta_{\ell m} = 1$ if $\ell = m$, and $\delta_{\ell m} = 0$ otherwise.

C. Sparse Approximations to the Transformed Pupil Mask

Figure 1 illustrates a circular pupil mask and its Fourier transform, which is an Airy pattern. The idea is to keep only a few points in the central core of the transformed mask to get a sparse approximation \tilde{M} to \tilde{M} .

The $n_x \times n_y$ array defined by the truncated mask generates a block-circulant circulant-block (BCCB) matrix.²¹ Figure 2(a) illustrates three different truncations of the transformed pupil mask, consisting of 1, 5, and 9 points in the central core. Here “o” denotes 1-point truncation, “o” plus “*” denotes 5-point truncation, and “o” plus “*” plus “□” denotes 9-point truncation. The sparsity patterns for the three corresponding BCCB matrices \tilde{M} are presented in Figs. 2(b)–2(d). In the case of 1-point truncation, \tilde{M} is diagonal. With 5-point truncation and 9-point truncation, \tilde{M} has five bands.

D. Sparse Matrix Reordering

Because the Fourier representer \hat{P} for the propagator has nonzero off-diagonal blocks, \hat{C} has a wide bandwidth. To reduce the bandwidth, and thus increase the efficiency of the solution to $\hat{C}\hat{z} = \hat{r}$, it is necessary to reorder the entries of \hat{C} .

An example of the representer \hat{C} of the truncated Fourier domain preconditioner with 1-point truncation is shown in Fig. 3(a), and the reordered \hat{C} is shown in Fig. 3(b). The reordered \hat{C} is a block diagonal matrix with $L \times L$ blocks. The cost of solving $\hat{C}\hat{z} = \hat{r}$ with this reordering is $\mathcal{O}(nL)$.

4. Combining Laser and Natural Guide Stars

A. Laser Guide Star Propagation Operators

Material in this section is similar to that in Ref. 19. Equations (13) and (14) model light propagation from a natural guide star to the ground level. In the case of laser guide stars, the propagation operators take a little different form because of the cone effect. If we still assume geometric optics propagation, then the

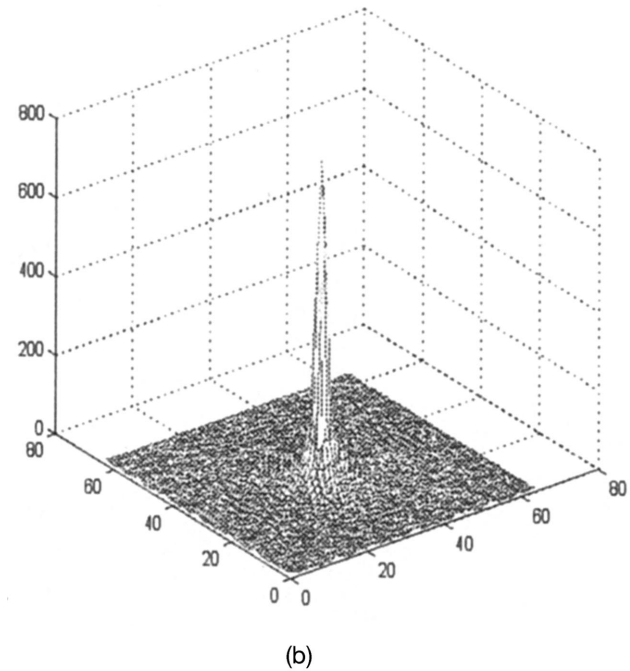
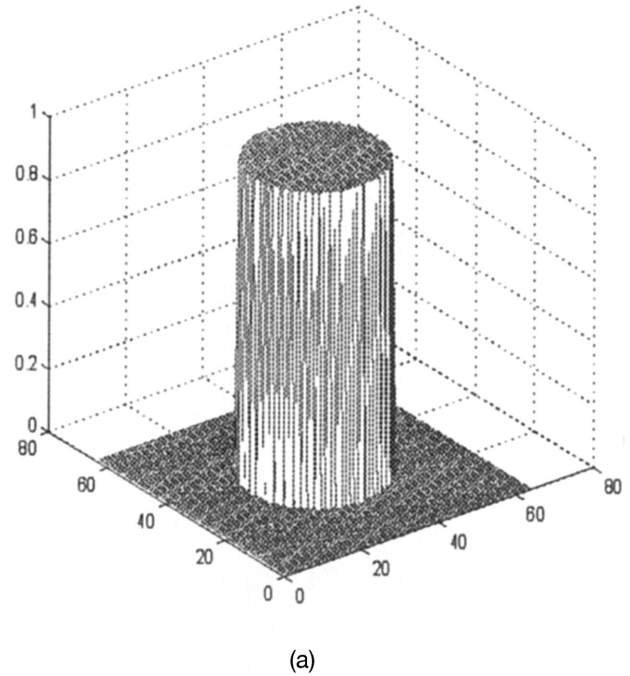


Fig. 1. (a) A circular pupil mask and (b) its Fourier transform.

pupil-plane phase that arises from propagation of light from a laser guide star in direction θ_j is represented as

$$\phi_j(\mathbf{x}) = [P^{\text{LGS}}(\theta_j)](\mathbf{x}) \stackrel{\text{def}}{=} \sum_{\ell=1}^L \psi_\ell \left(\frac{H}{H - z_\ell} \mathbf{x} + z_\ell \theta_j \right), \quad (25)$$

where H denotes the altitude of the laser guide star.

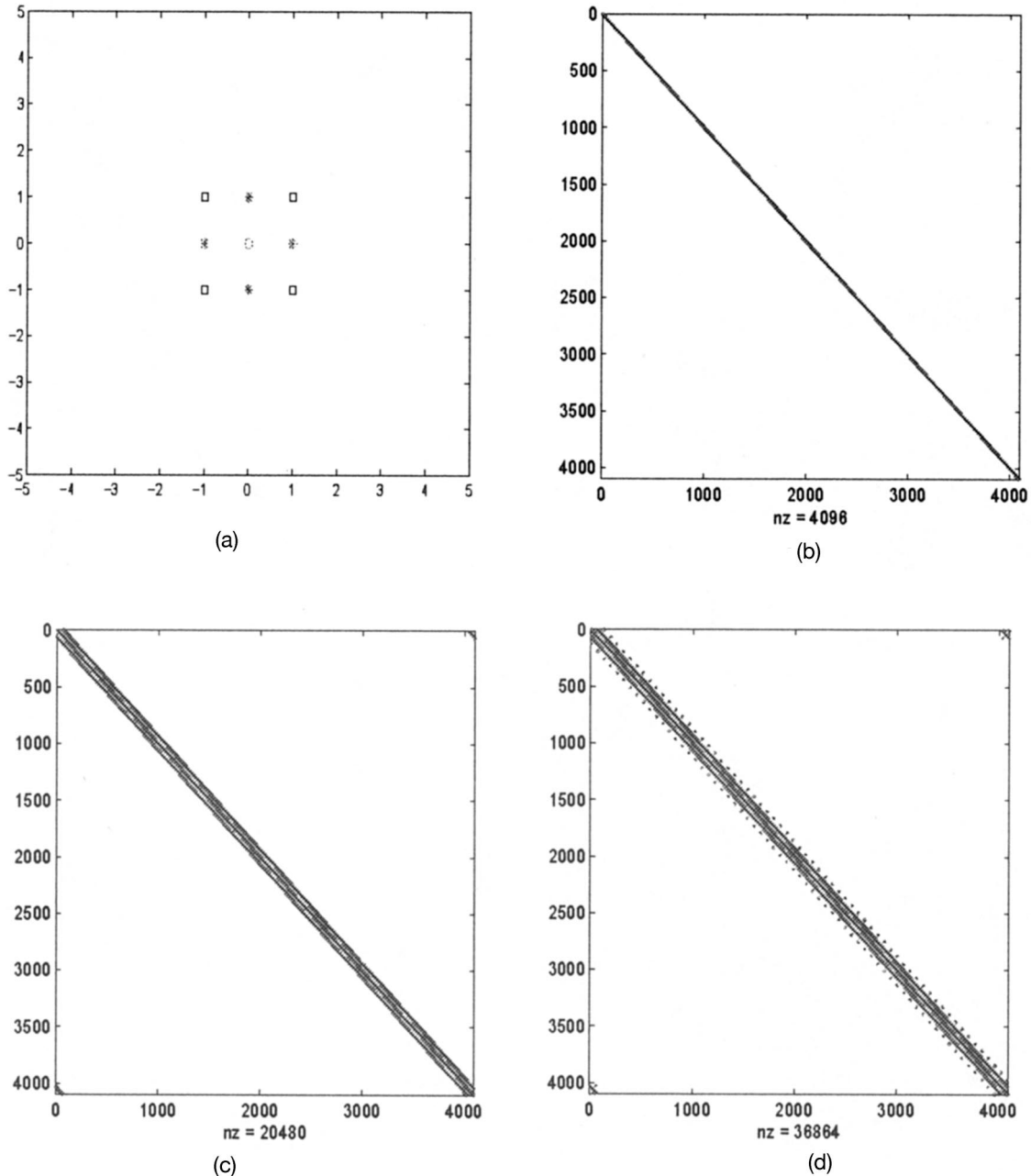


Fig. 2. Truncations of the transformed pupil mask and the sparsity patterns of the corresponding BCCB matrix approximations \tilde{M} to \tilde{M} . (a) Shows the truncation points; subplot (b) shows the sparsity pattern of the BCCB matrix \tilde{M} corresponding to 1-point truncation; (c) corresponds to 5-point truncation; and (d) corresponds to 9-point truncation.

The Fourier transform of Eq. (25) is then

$$\hat{\phi}_j(\kappa) = \sum_{\ell=1}^L \frac{1}{c_\ell} \exp[2\pi i(\kappa/c_\ell)(z_\ell \theta_j)] \hat{\psi}_\ell(\kappa/c_\ell), \quad (26)$$

where $\hat{\psi}_\ell = \mathcal{F}(\psi_\ell)$ and

$$c_\ell = \frac{H}{H - z_\ell}. \quad (27)$$

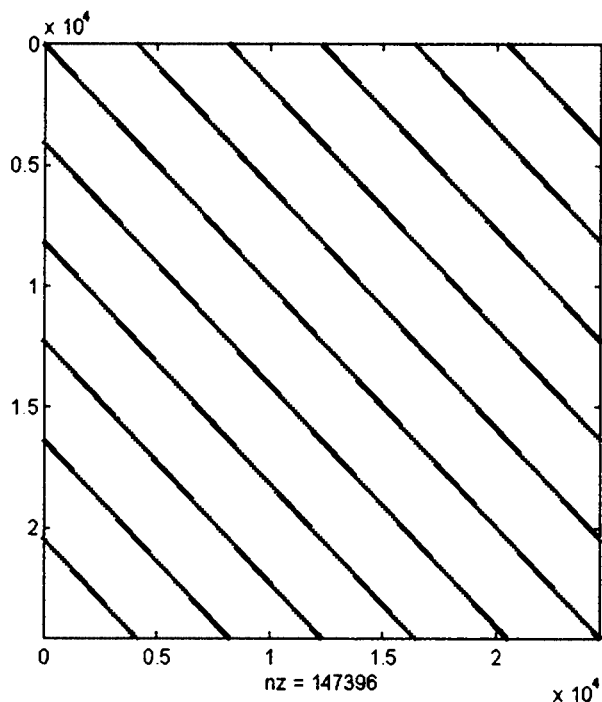
From Eq. (26) we see that the spatial frequency κ is

squeezed by a factor of c_ℓ when light propagates from altitude z_ℓ to the ground level.

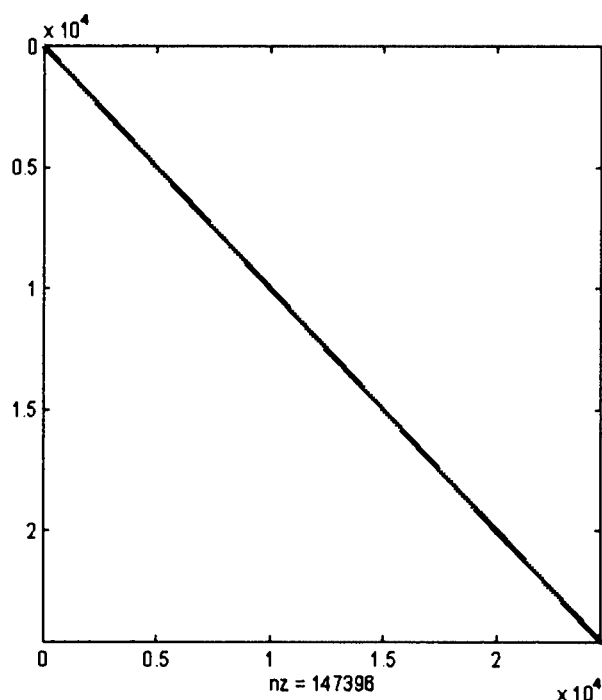
In order to avoid interpolation, we allow the grid spacing to vary with layer height z_ℓ . In particular, we take the spacing to inversely proportional to c_ℓ .

B. Low-Rank Perturbations to the Noise Covariance Matrix

The noise covariance matrix C_η in Eq. (20) is no longer diagonal when position uncertainty of a laser guide star is considered. The exact position of a LGS projected into the sky is variable as a result of both



(a)



(b)

Fig. 3. Sparsity pattern for the representer \hat{C} of the Fourier domain preconditioner with 1-point truncation. (a) Shows the sparsity pattern without reordering, and (b) is with reordering.

fundamental effects of atmospheric turbulence and practical error sources in the laser system. At least for now, there is no independent means of measuring

the actual position with any accuracy. For this problem, one or more natural guide stars must be included in the guide star constellation of a LGS AO system to estimate the global tip-tilt mode of the wavefront sensor.

We follow the model proposed by Ellerbroek²⁰ for LGS position uncertainty. In the discussion to follow, n_{lgs} denotes the number of laser guide stars and n_{tt} denotes the number of natural tip-tilt guide stars. WFS measurements defined by Eq. (1) now are split into two parts: a higher-order component s_h and a tip-tilt component s_t , so

$$s = \begin{bmatrix} s_h \\ s_t \end{bmatrix} = \begin{bmatrix} G_h \\ G_t \end{bmatrix} \psi + \begin{bmatrix} \eta_h \\ \eta_t \end{bmatrix}. \quad (28)$$

Here G_t is a low-rank matrix whose number of rows equals $2n_{\text{tt}}$, and η_h and η_t represent noise for higher-order sensor signals and tip-tilt sensor signals, respectively.

C_η takes the form

$$C_\eta = \begin{bmatrix} N_h + \sigma_t^2 T T^T & 0 \\ 0 & N_t \end{bmatrix}, \quad (29)$$

where N_h and N_t are diagonal matrices that describe the statistics of the noise within the higher-order and tip-tilt WFS themselves. T is a low-rank matrix with $2n_{\text{lgs}} \times 2n_{\text{lgs}}$ block structure, and each block is a column vector of ones of length equal to the number of sensor subapertures within the pupil. The scalar σ_t is the rms position uncertainty for each LGS.

As $\sigma_t^2 \rightarrow \infty$ we obtain

$$C_\eta^{-1} = \begin{bmatrix} N_h^{-1/2} (I - Q) N_h^{-1/2} & 0 \\ 0 & N_t^{-1} \end{bmatrix}, \quad (30)$$

where Q is the noise-weighted orthogonal projector onto the range of T ,

$$Q = \tilde{T} (\tilde{T}^T \tilde{T})^{-1} \tilde{T}^T, \quad \text{with } \tilde{T} = N_h^{-1/2} T. \quad (31)$$

We next combine Eqs. (6) and (30) with the decomposition of G in Eq. (28), and we set

$$\tilde{G}_h = N_h^{-1/2} G_h, \quad \tilde{G}_t = N_t^{-1/2} G_t \quad (32)$$

to obtain

$$\begin{aligned} A &= \tilde{G}_h^T (I - Q) \tilde{G}_h + \tilde{G}_t^T \tilde{G}_t + C_\psi^{-1} \\ &= A_h - (\tilde{G}_h^T Q \tilde{G}_h - \tilde{G}_t^T \tilde{G}_t), \end{aligned} \quad (33)$$

where

$$A_h = G_h^T N_h^{-1} G_h + C_\psi^{-1}. \quad (34)$$

We can express the second term on the right-hand

Table 1. Atmospheric Turbulence Profile Used for Simulations

Layer	Altitude (km)	Relative Layer Weight
1	0.00	0.652
2	2.58	0.172
3	5.16	0.055
4	7.73	0.025
5	12.89	0.074
6	15.46	0.022

side of Eq. (33) as UV^T , where $U = [U_1 U_2]$ and $V = [U_1 - U_2]$, with

$$U_2 = \tilde{G}_t^T = G_t^T N_t^{-1/2}, \quad (35)$$

$$U_1 = \tilde{G}_h^T \tilde{T} R = G_h^T N_h^{-1} T R. \quad (36)$$

To derive U_1 we used Eqs. (31) and (32), and we selected R so that $RR^T = (\tilde{T}^T \tilde{T})^{-1} = (T^T N_h T)^{-1}$. Such an R can be computed using the Choleski factorization. By applying the Sherman–Morrison formula²³ to Eq. (33), we obtain

$$A^{-1} = A_h^{-1} + L, \quad (37)$$

where

$$\begin{aligned} L &= (A_h^{-1} U) (I - V^T A_h^{-1} U)^{-1} (A_h^{-1} V)^T \\ &= [A_h^{-1} U_1 \quad A_h^{-1} G_t^T] \begin{bmatrix} I_1 - U_1^T A_h^{-1} U_1 & -U_1^T A_h^{-1} G_t^T \\ G_t A_h^{-1} U_1 & N_t + G_t A_h^{-1} G_t^T \end{bmatrix}^{-1} \begin{bmatrix} (A_h^{-1} U_1)^T \\ -(A_h^{-1} G_t^T)^T \end{bmatrix}. \end{aligned} \quad (38)$$

From Eq. (37),

$$A^{-1} b = A_h^{-1} b + L b. \quad (39)$$

To efficiently perform this computation, we first precompute $W_1 = A_h^{-1} U_1$ and $W_2 = A_h^{-1} G_t^T$. This can be done by applying FD-PCG to solve matrix-vector systems $A_h w = u$ for each of the $2n_{\text{igs}}$ columns of U_1 and each of the $2n_{\text{tt}}$ columns of G_t^T . We then precompute the inner matrix inverse in Eq. (38); call it Z . Note that Z is relatively small, with $2(n_{\text{igs}} + n_{\text{tt}})$ rows and columns. The matrix-vector product Lb is easy to compute once W_1 , W_2 , and Z are available. We also compute $A_h^{-1} b$ using FD-PCG. The preconditioner is nearly identical to that introduced in Section 3; the

only significant difference is in the propagation operators; compare Eq. (13) with Eq. (25).

5. Numerical Simulation Results

A. Cases Considered and Performance Evaluation Criteria

Table 1 presents the atmospheric turbulence profile used for the simulations described in this section. This six-layer profile is based upon thermosonde and generalized scintillation detection and ranging measurements taken at Cerro Pachon, Chile.

Figures 4(a) and 4(b) present the configuration, or constellation, of guide stars for NGS and for LGS plus NGS, respectively. In the latter case, the NGSs are used to estimate global tip-tilt. The dashed lines in these figures indicate science directions from $[-0.5, 0]$ arc min to $[+0.5, 0]$ arc min over which phase errors will be evaluated.

Results for two different computational grids will be presented. By the 64×64 grid we mean that a 64×64 grid was used for each of each of the six turbulence layers, and similarly for the 128×128 grid. In either case, the ground-layer grid covers a square computational domain whose width is twice the diameter of the telescope.

Let x_k denote approximation to the solution of $Ax = b$ at iteration k , and let ψ_k denote corresponding estimate for the atmospheric turbulence profile. We measure AO performance by the (pupil-plane integrated residual squared) phase error,

$$\text{PE}_k(\theta) = \int_{\Omega} \{ [P(\theta) \psi_k](\mathbf{x}) - (P(\theta) \psi_{\text{atmos}})(\mathbf{x}) \}^2 d\mathbf{x}. \quad (40)$$

Here $P(\theta)$ represents natural guide star propagation in direction θ in the field of view [see Eq. (13)] and Ω represents the pupil plane. By integrating over the science field of view, we obtain the FOV-integrated phase error,

$$\text{FOVPE}(k) = \int_{\text{FOV}} \text{PE}_k(\theta) d\theta. \quad (41)$$

In the comparisons to follow, we will display normalized rms phase errors and normalized FOV-

integrated phase errors, given by the square roots of ratios,

$$\sqrt{\frac{\text{PE}_k(\theta)}{\int_{\Omega} [P(\theta)\psi_{\text{atmos}}]^2(\mathbf{x})d\mathbf{x}}},$$

$$\sqrt{\frac{\text{FOVPE}(k)}{\int_{\text{FOV}} \int_{\Omega} [P(\theta)\psi_{\text{atmos}}]^2(\mathbf{x})d\mathbf{x}d\theta}},$$

respectively. Each error curve will represent an average over five turbulence profile and sensor noise realizations. The above denominators represented incident phase aberrations integrated over the telescope aperture, and integrated over aperture and FOV, respectively. These ratios can be interpreted as inverse correction factors. The correction factors are typically around 20. For a 30 m telescope with $r_0 = 0.16$ m, a typical incident phase aberration on the pupil plane in our simulations is around 4 μm rms. Thus the residual phase error is around 200 nm rms.

The turbulence layers are generated using a discrete Fourier domain Monte Carlo approach,²⁴ with the Kolmogorov power spectrum scaled according to turbulence strength; see Eq. (11). The width of the computational domain for the turbulence layers is taken to be twice the telescope diameter. Because of the periodicity of the discrete Fourier transform, this produces an artificial outer scale comparable to the telescope diameter.

The multigrid method used in our comparisons requires a sparse approximation to the covariance matrix C in Eq. (11). Following Ellerbroek²⁰ we take an approximation of the form $\gamma^2 L^{-2}$, where L is a discretized Laplacian operator. The scaling parameter is chosen to be

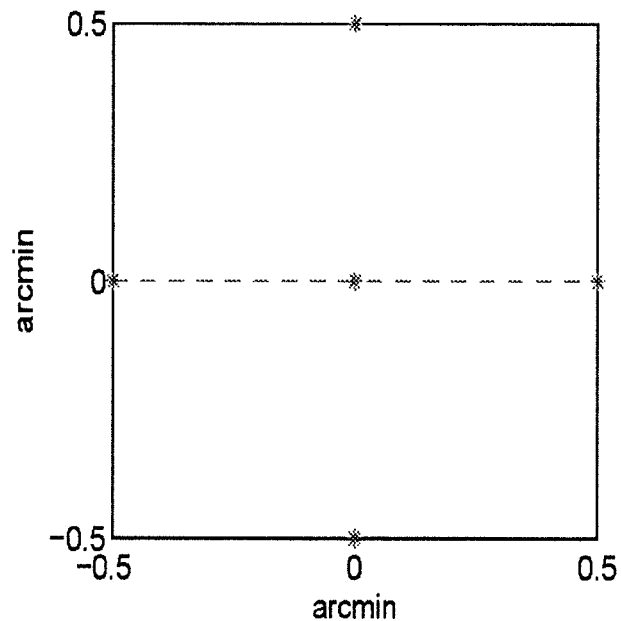
$$\gamma = \text{trace}(L^{-1}C^{1/2})/\text{trace}(L^{-2}). \quad (42)$$

This choice minimizes $\langle \|C^{1/2}w - \gamma L^{-1}w\|^2 \rangle$, where w denotes discrete white noise. Both the numerator and denominator in Eq. (42) can be efficiently evaluated using a Monte Carlo scheme²¹ known as randomized trace.

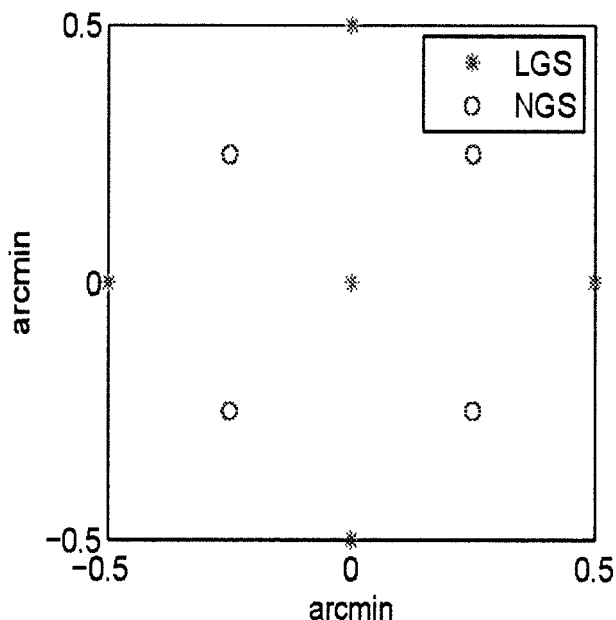
We quantify the amount of noise in the WFS gradient measurements by signal-to-noise ratio (SNR),

$$\text{SNR} = \sqrt{\frac{\text{trace}(GC_{\psi}G^T)}{2\sigma^2 N_{\text{gs}} N_{\text{sensor}}}}, \quad (43)$$

where N_{gs} is the number of guide stars and N_{sensor} is the number of sensor subapertures. This is different from the standard definition of SNR for photometry,



(a)



(b)

Fig. 4. Guide star constellations. (a) NGS and (b) LGS with NGSs added in to measure global tip-tilt. Phase errors will be evaluated along the dashed lines.

$$\text{SNR} = \frac{\bar{N}}{\sqrt{N + \sigma^2}}, \quad (44)$$

where \bar{N} is the expected number of photoelectrons falling on each sensor element and σ is the read noise. Equation (43) relates more directly to the performance of a wavefront reconstruction algorithm, while Eq. (44)

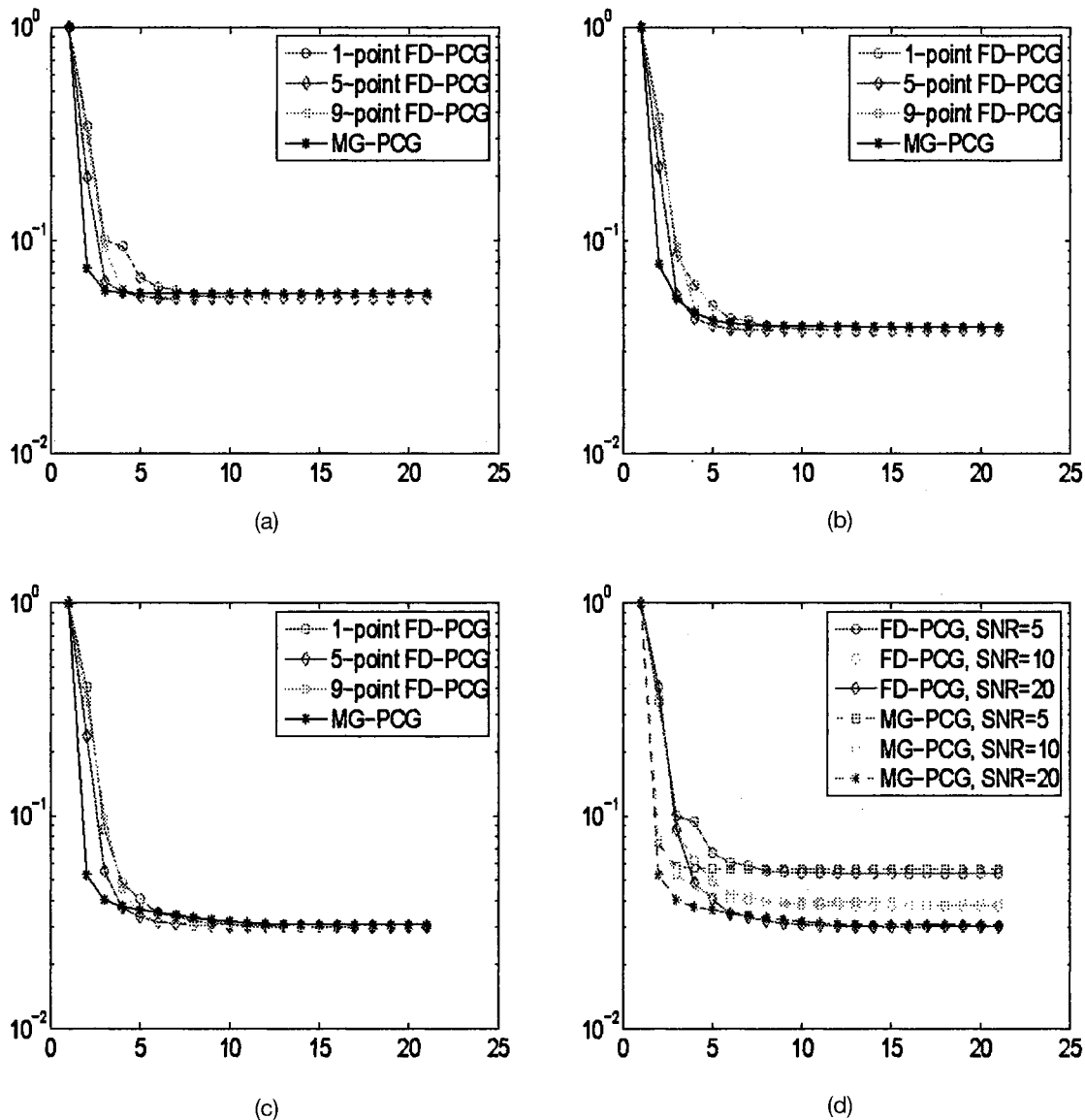


Fig. 5. Normalized rms FOV-integrated phase errors for MG-PCG versus FD-PCG, with natural guide stars. In (a) we vary the FD-PCG truncation level with $\text{SNR} = 5$; in (b) we vary the truncation level with $\text{SNR} = 10$; and in (c) we vary the truncation level with $\text{SNR} = 20$. In (d) 1-point truncation is used for FD-PCG, and the noise level is varied. The grid size is 64×64 and horizontal axes give iteration count k .

cannot be used to predict wavefront reconstruction accuracy without detailed knowledge of the sensor model, e.g., the spot size for a Shack–Hartmann sensor.

By truncation level, we mean the number of points in the central core of the transformed pupil mask that we keep to construct the sparse approximation \tilde{M} to the masking operator \tilde{M} (see Subsection 3.C.).

Finally, we note that the operator G_i in Eq. (28) is based on an rms-best-fit tilt model of the tip–tilt sensor, and not on an average phase gradient model.

B. Natural Guide Star Tomography

FOV-integrated phase errors [see Eq. (41)] are compared in Fig. 5. In all cases we see that the phase errors reach their asymptotic limit, or “plateau,” in

approximately 10 iterations. This plateau occurs at earlier iterations as the error level increases. At all noise levels, increasing the truncation for FD-PCG increases the convergence rate, but the advantage of taking more points in the truncated pupil mask appears to decrease at lower noise levels. At the lowest noise level ($\text{SNR} = 20$) FD-PCG converges faster than MG-PCG.

In Fig. 6 we examine the behavior of the phase error as a function of view angle θ . Here θ ranges from $[-0.5, 0]$ arc min to $[+0.5, 0]$ arc min. This is represented by the dashed line in Fig. 4(a). From this figure we see that for both preconditioners, residual errors reach their asymptotic limit after about six PCG iterations. The residual errors in the directions of guide stars are only 1/5 of those at

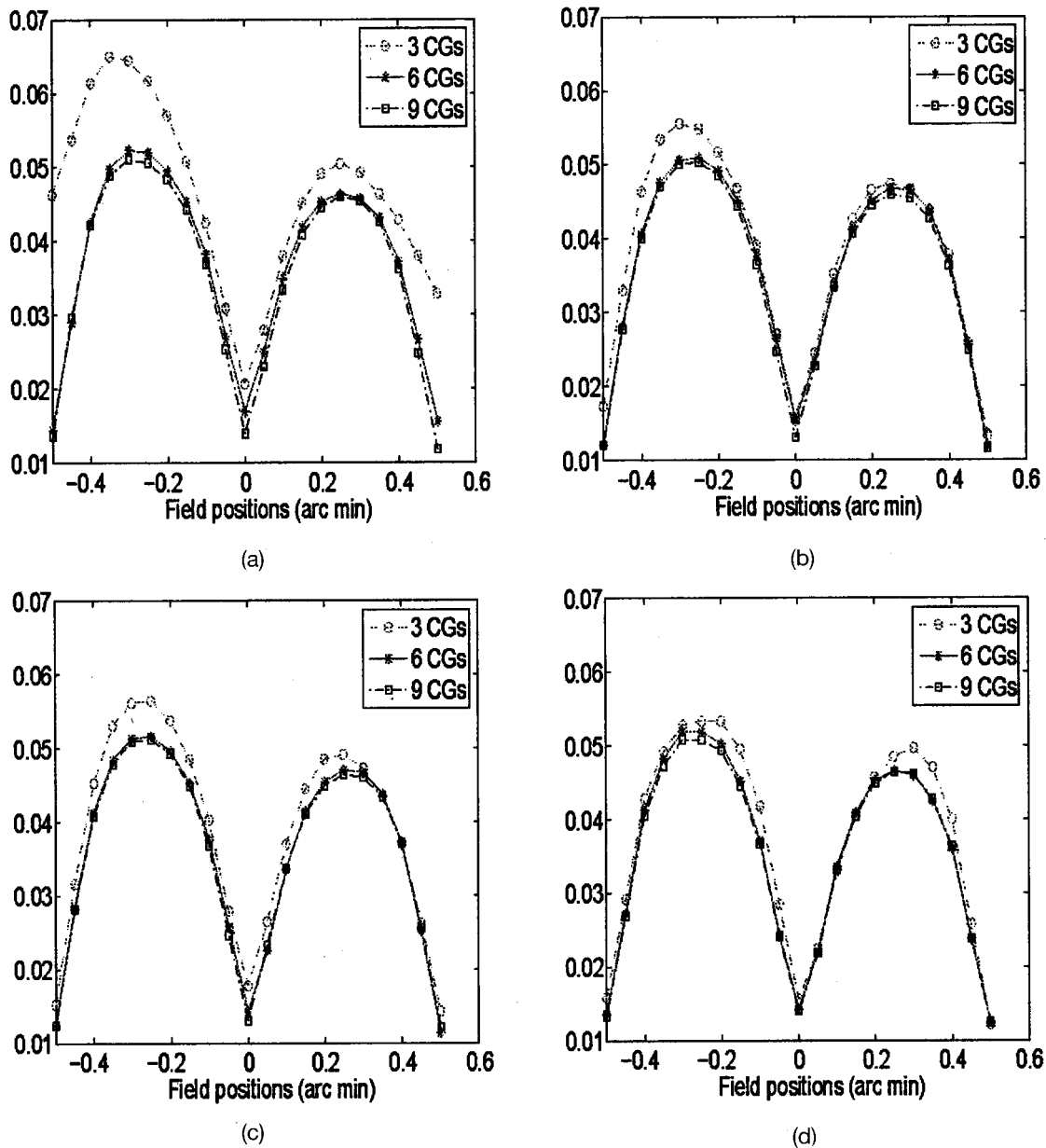


Fig. 6. Normalized rms phase errors when the science direction θ is sampled along an arc across the field of view. (a) 1-point FD-PCG, (b) 5-point FD-PCG, (c) 9-point FD-PCG, and (d) MG-PCG.

the midpoints between two neighboring guide stars.

Figure 7 shows the effects of changing the grid resolution on PCG convergence rates. From this figure we see that MG-PCG convergence is again sensitive to the system size, but the FD-PCG convergence is not. At grid size 64×64 , the MG-PCG needs fewer iterations to converge than FD-PCG does, but at grid size 128×128 , the two preconditioners need almost the same number of iterations to reach their asymptotic limits.

In Table 2 we compare computational time per iteration for MG-PCG versus FD-PCG. Results are shown for various truncation levels with grid size 64×64 and for 1-point truncation with a 128×128 grid. We see that for all truncation levels, preconditioning time

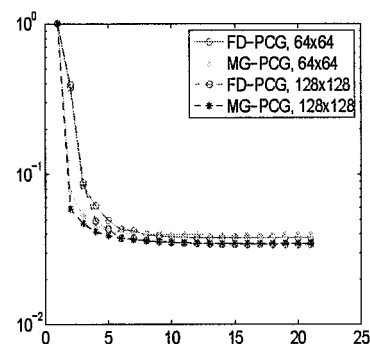


Fig. 7. Grid resolution effects on PCG convergence, with natural guide stars. Normalized FOV-averaged phase errors are compared for MG-PCG and FD-PCG with 1-point truncation. Grid sizes are 64×64 and 128×128 and $\text{SNR} = 10$.

Table 2. Computation Time per Iteration for MG-PCG versus FD-PCG for Tomography with Five Natural Guide Stars^a

Grid size	Preconditioner	Total (s)	C^{-1} Mult (s)	A Mult (s)
64×64	MG-PCG	13.66	13.60	0.04
	9-point FD-PCG	4.25	4.01	0.21
	5-point FD-PCG	2.49	2.25	0.21
	1-point FD-PCG	0.32	0.07	0.21
128×128	MG-PCG	75.60	75.42	0.20
	1-point FD-PCG	1.26	0.31	0.88

^aThe column “Total” denotes the total time per PCG iteration; column “ C^{-1} mult” denotes the preconditioning time, and column “A mult” denotes the time for matrix-vector multiplication by A .

per FD-PCG iteration is much less than the preconditioning time for MG-PCG. With 1-point truncation, the preconditioning times differ by more than two orders of magnitude. The cost of multiplication by A is much less for MG-PCG than for FD-PCG. This is due to the fact that we use a sparse matrix propagator for MG-PCG and a Fourier propagator for FD-PCG. Overall MG-PCG computation times are dominated by the preconditioning, so FD-PCG is much faster than MG-PCG. Since total FD-PCG time per iteration scales as $n \log n$, while total MG-PCG time scales as $n^{3/2}$, as the grid size increases the ratio of FD-PCG time relative to MG-PCG should become smaller.

The total time required to solve the system $Ax = b$ is equal to the time per iteration multiplied by the number of iterations. As the noise level decreases, we saw in Fig. 7 that the number of iterations needed to attain asymptotic FOV-integrated phase error increases significantly for MG-PCG, but it increases only slightly for FD-PCG. In particular, for SNR = 5 on a 64×64 grid, MG-PCG needs 2–3 iterations, and FD-PCG with 1-point truncation needs 7–8 iterations, so the total computation time for MG-PCG is 10 times the total computation times for FD-PCG. On the other hand for SNR = 20 on a 128×128 grid, MG-PCG and FD-PCG with 1-point truncation both need about 9 iterations, so there is a factor of 60 between the total computational times.

C. Laser Guide Stars with Position Uncertainty

In this subsection we examine a more realistic tomography system in which five laser guide stars are used to estimate higher-order wavefront aberrations. Because of the position uncertainty of the laser guide stars, four natural guide stars are added to estimate global tip and tilt.

Figure 8 shows the phase errors for LGS + NGS with these two preconditioners for various noise levels. From this figure we see that the dependence of PCG convergence on the SNR in the LGS + NGS case is nearly identical to the dependence in the LGS case. We also found almost no difference between LGS + NGS and NGS when we examined convergence dependence on grid size. Finally, we found the timing results for LGS + NGS to be almost identical to those

shown in Table 2 for NGS. This is because the cost of the low-rank matrix-vector product Lb in Eq. (39) is small relative to the cost of computing $A_h^{-1}b$.

Because of tip-tilt uncertainty, the accuracy of LGS + NGS is a little worse than that of NGS alone. Figure 9 shows that at moderate noise levels, the FOV-integrated phase error for NGS is about 14% smaller than the error for LGS + NGS.

6. Summary

In this paper we introduced a Fourier domain preconditioned conjugate gradient (FD-PCG) method for atmospheric tomography. We presented results of a numerical study in which we compared FD-PCG against an existing multigrid preconditioned conjugate gradient (MG-PCG) method. This study involved two types of illumination to probe the atmosphere: (i) natural guide stars, and (ii) laser guide stars in combination with natural guide stars to estimate tip-tilt. From this study we draw the following conclusions:

(A) The FD-PCG computation time per iteration is much less than for MG-PCG. For grid sizes required to accurately model a 30 m telescope, we found FD-PCG to be about 60 times faster per iteration than MG-PCG. The asymptotic cost per iteration of FD-PCG is $\mathcal{O}(n \log n)$, while the cost per iteration of MG-PCG is $\mathcal{O}(n^{3/2})$. Therefore the advantage of FD-PCG becomes more significant when system size increases. FD-PCG also requires significantly less memory than does MG-PCG.

(B) For realistic sensor noise levels, FD-PCG and MG-PCG both converge in approximately 10 iterations. FD-PCG convergence rates are relatively stable with respect to changes in noise level. On the other hand, MG-PCG requires fewer iterations at high noise levels and more iterations at low noise levels. From item A we then conclude that the overall computation time for FD-PCG is from one to two orders of magnitude less than overall MG-PCG computation time.

(C) In our simulations, almost the same residual pupil-plane phase errors are obtained from FD-PCG and MG-PCG. MG-PCG errors are slightly larger than those from FD-PCG, due to the sparse biharmonic, or “curvature,” operator approximation to the covariance obtained from a Kolmogorov turbulence model.

The following conclusions seem reasonable, but were not validated in this paper:

(D) By making minor modifications to the pupil mask operator, one can incorporate subsampling. By this we mean that the grid spacing in the computational domain can be made much smaller than the distance between sensor elements. This is important for accurately modeling the sensors and in determining the fitting error. With the modifications required for subsampling, the sparsity structure of trans-

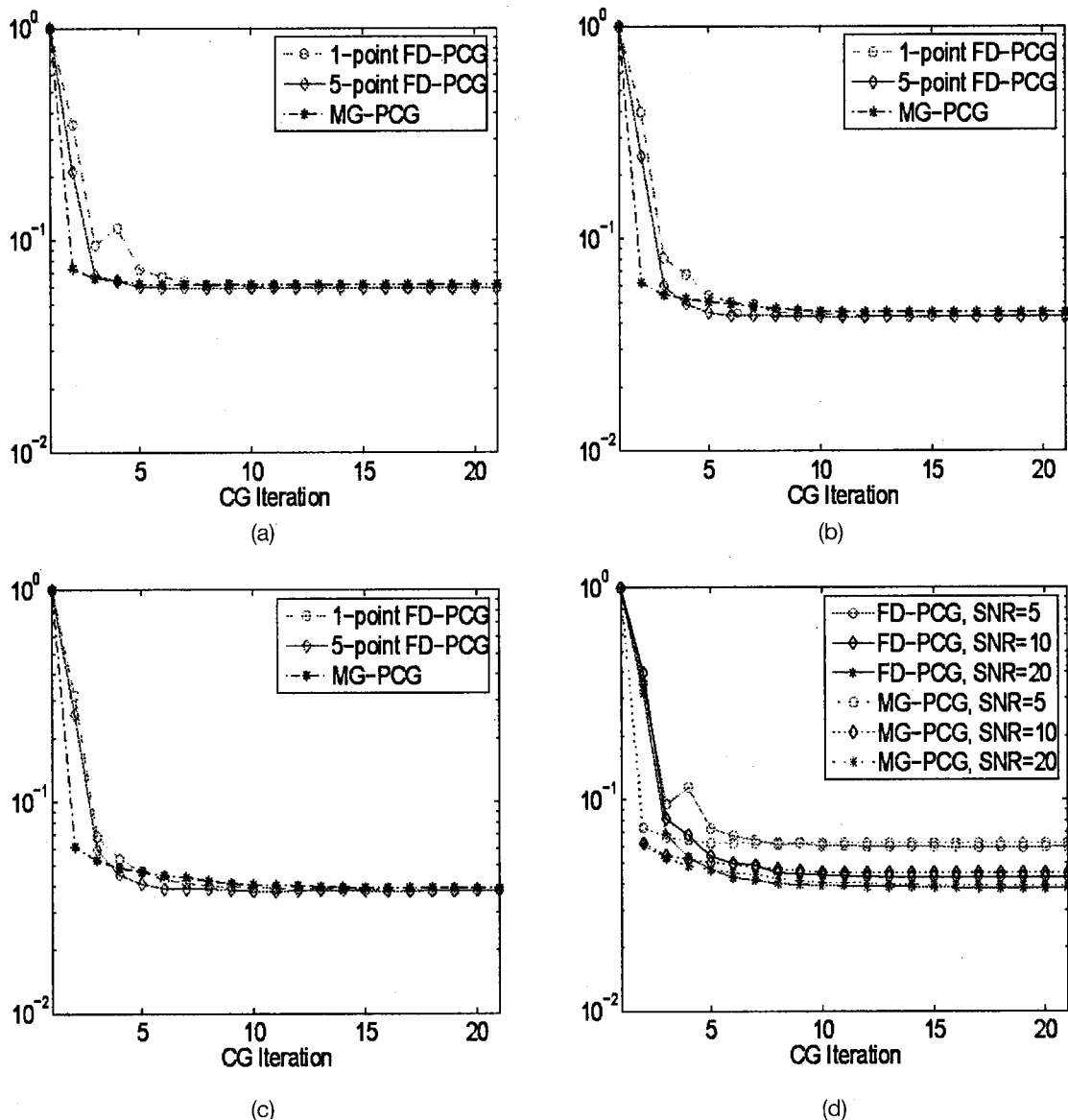


Fig. 8. Normalized rms FOV-integrated phase errors for MG-PCG versus FD-PCG, with five laser guide stars and four natural guide stars. (a) SNR = 5, (b) SNR = 10, (c) SNR = 20, and (d) compares MG-PCG and 1-point FD-PCG under three different noise levels. 64×64 grid size.

formed pupil mask \hat{M} does not significantly change, nor is the computational cost significantly increased.

(E) Computation time for FD-PCG can be significantly decreased by using special-purpose hardware to compute the fast Fourier transform (FFT), and by implementing certain parts of the FD-PCG algorithm in parallel. In particular, with 1-point truncation of the transformed pupil mask and with sparse matrix reordering, the Fourier representer \hat{C} for the preconditioner is sparse with very small bandwidth. Thus most of the computational cost of the preconditioner involves applying 2D FFTs to arrays that have a block decomposition associated with the layered structure of the atmosphere. FFTs can easily be applied in parallel to these blocks.

(F) Additional savings can be achieved by varying

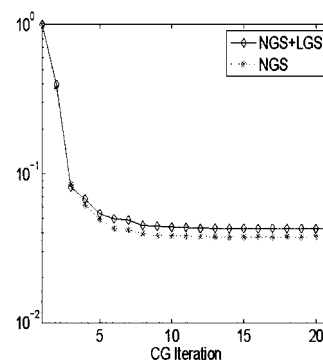


Fig. 9. Normalized rms FOV-integrated phase errors for LGS + NGS versus NGS. Systems $Ax = b$ were solved using FD-PCG with 1-point truncation. The SNR = 10 and the grid size is 64×64 .

the grid spacing with layer height. In particular, at higher altitudes the turbulence is weaker, so larger grid spacing (fewer grid points) can be used without significantly decreasing accuracy.

This work has been supported in part by the National Science Foundation Science and Technology Center for Adaptive Optics, managed by the University of California at Santa Cruz under cooperative agreement AST-9876783. Q. Yang and C. R. Vogel received additional partial support from the Computational Mathematics Program at the Air Force Office of Scientific Research through AOSR-DEPSCoR grant F49620-02-1-0297.

B. L. Ellerbroek received support from the Thirty-Meter Telescope (TMT) project. TMT is a partnership of the Association of Universities for Research in Astronomy (AURA), the Association of Canadian Universities for Research in Astronomy (ACURA), the California Institute of Technology, and the University of California. The partners gratefully acknowledge the support of the Gordon and Betty Moore Foundation, the U.S. National Science Foundation, the National Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada, and the Gemini Partnership.

References

1. R. G. Dekany, M. C. Britton, D. T. Gavel, B. L. Ellerbroek, G. Herriot, C. E. Max, and J.-P. Veran, "Adaptive optics requirements definition for TMT," in *Advancements in Adaptive Optics*, D. B. Calia, B. L. Ellerbroek, and R. Ragazzoni, eds., Proc. SPIE **5490**, 879–890 (2004).
2. T. E. Andersen, A. Ardeberg, H. Riewaldt, M. Lastiwka, N. Quinlan, K. McNamara, X. Wang, A. Enmark, M. Owner-Petersen, A. Shearer, C. Fan, and D. Moraru, "Status of the Euro50 Project," in *Ground-Based Telescopes*, J. M. Oschmann, Jr., ed., Proc. SPIE **5489**, 407–416 (2004).
3. M. Johns, J. R. P. Angel, S. Shectman, R. Bernstein, D. G. Fabricant, P. McCarthy, and M. Phillips, "Status of the Giant Magellan Telescope (GMT) Project," in *Ground-Based Telescopes*, J. M. Oschmann, Jr., ed., Proc. SPIE **5489**, 441–453 (2004).
4. P. Dierickx, E. T. Brunetto, F. Comeron, R. Gilmozzi, F. Y. Gonté, F. Koch, M. le Louarn, G. J. Monnet, J. Spyromilio, I. Surdej, C. Verinaud, and N. Yaitskova, "OWL phase A status report," in *Ground-Based Telescopes*, J. M. Oschmann, Jr., ed., Proc. SPIE **5489**, 391–406 (2004).
5. J. M. Beckers, "Increasing the size of the isoplanatic patch with multi-conjugate adaptive optics," in *Proceedings of European Southern Observatory Conference and Workshop on Very Large Telescopes and Their Instrumentation*, M. H. Ulrich, ed. (European Southern Observatory, 1988), Vol. 30, pp. 693–703.
6. D. C. Johnston and B. M. Welsh, "Analysis of multiconjugate adaptive optics," J. Opt. Soc. Am. A **11**, 394–408 (1994).
7. B. L. Ellerbroek, "First-order performance evaluation of adaptive-optics systems for atmospheric turbulence compensation in extended field-of-view astronomical telescopes," J. Opt. Soc. Am. A **11**, 783–805 (1994).
8. T. Fusco, J. M. Conan, G. Rousset, L. M. Mugnier, and V. Michau, "Optimal wave-front reconstruction strategies for multiconjugate adaptive optics," J. Opt. Soc. Am. A **18**, 2527–2538 (2001).
9. A. Tokovinin and M. Le Louarn, "Isoplanatism in a multiconjugate adaptive optics system," J. Opt. Soc. Am. A **17**, 1819–1827 (2000).
10. A. Tokovinin and E. Viard, "Limiting precision of tomographic phase estimation," J. Opt. Soc. Am. A **18**, 873–8827 (2001).
11. B. L. Ellerbroek, "Linear systems modeling of adaptive optics in the spatial-frequency domain," J. Opt. Soc. Am. A **22**, 310–322 (2005).
12. E. P. Wallner, "Optimal wave-front correction using slope measurement," J. Opt. Soc. Am. **73**, 1771–1776 (1983).
13. B. M. Welsh and C. S. Gardner, "Performance analysis of adaptive-optics systems using laser guide stars and slope sensors," J. Opt. Soc. Am. A **6**, 1913–1923 (1989).
14. L. Gilles, C. R. Vogel, and B. L. Ellerbroek, "Multigrid preconditioned conjugate-gradient method for large-scale wave-front reconstruction," J. Opt. Soc. Am. A **19**, 1817–1822 (2002).
15. L. Gilles, B. L. Ellerbroek, and C. R. Vogel, "Preconditioned conjugate gradient wavefront reconstructors for multiconjugate adaptive optics," Appl. Opt. **42**, 5233–5250 (2003).
16. B. L. Ellerbroek, L. Gilles, and C. R. Vogel, "Numerical simulations of multiconjugate adaptive optics wavefront reconstruction on giant telescopes," Appl. Opt. **42**, 4811–4818 (2003).
17. H. Ren and R. Dekany, "Fast wave-front reconstruction by solving the Sylvester equation with the alternating direction implicit method," Opt. Express **12**, 3279–3296 (2004).
18. L. A. Poyneer, D. T. Gavel, and J. M. Brase, "Fast wave-front reconstruction in large adaptive optics systems with use of the Fourier transform," J. Opt. Soc. Am. A **19**, 2100–2111 (2002).
19. D. Gavel, "Tomography for multiconjugate adaptive optics systems using laser guide stars," in *Advancements in Adaptive Optics*, D. Bonaccini Calia, B. Ellerbroek, and R. Ragazzoni, eds., Proc. SPIE **5490**, 1356–1373 (2004).
20. B. L. Ellerbroek, "Efficient computation of minimum-variance wave-front reconstructors with sparse matrix techniques," J. Opt. Soc. Am. A **19**, 1803–1816 (2002).
21. C. R. Vogel, *Computational Methods for Inverse Problems* (SIAM, 2002).
22. Y. Saad, *Iterative Methods for Sparse Linear System*, 2nd ed. (SIAM, 2003), pp. 261–282.
23. G. H. Golub and C. F. VanLoan, *Matrix Computations*, 2nd ed. (Johns Hopkins U. Press, 1989).
24. M. C. Roggemann and B. Welsh, *Imaging Through Turbulence* (CRC 1996).